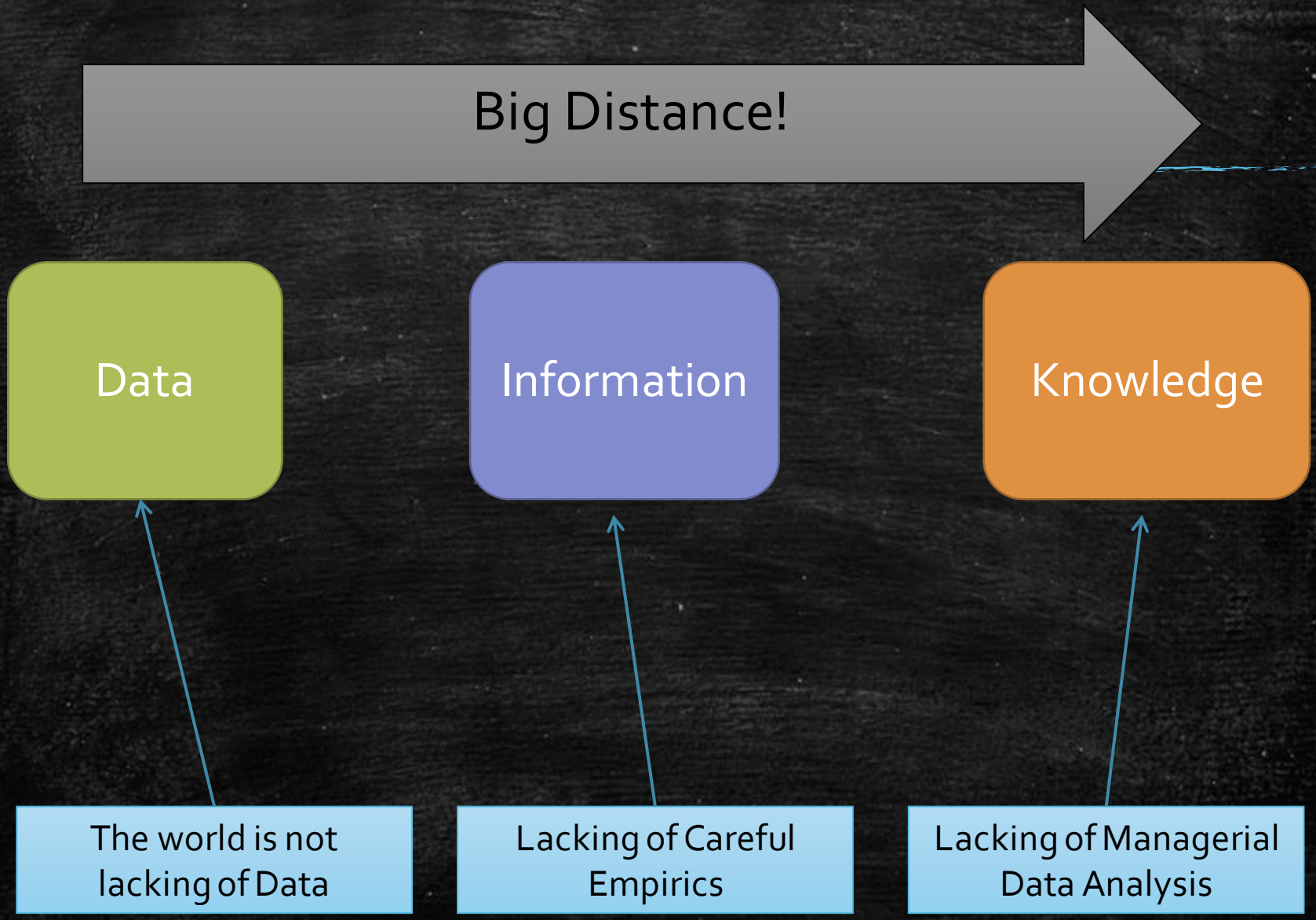
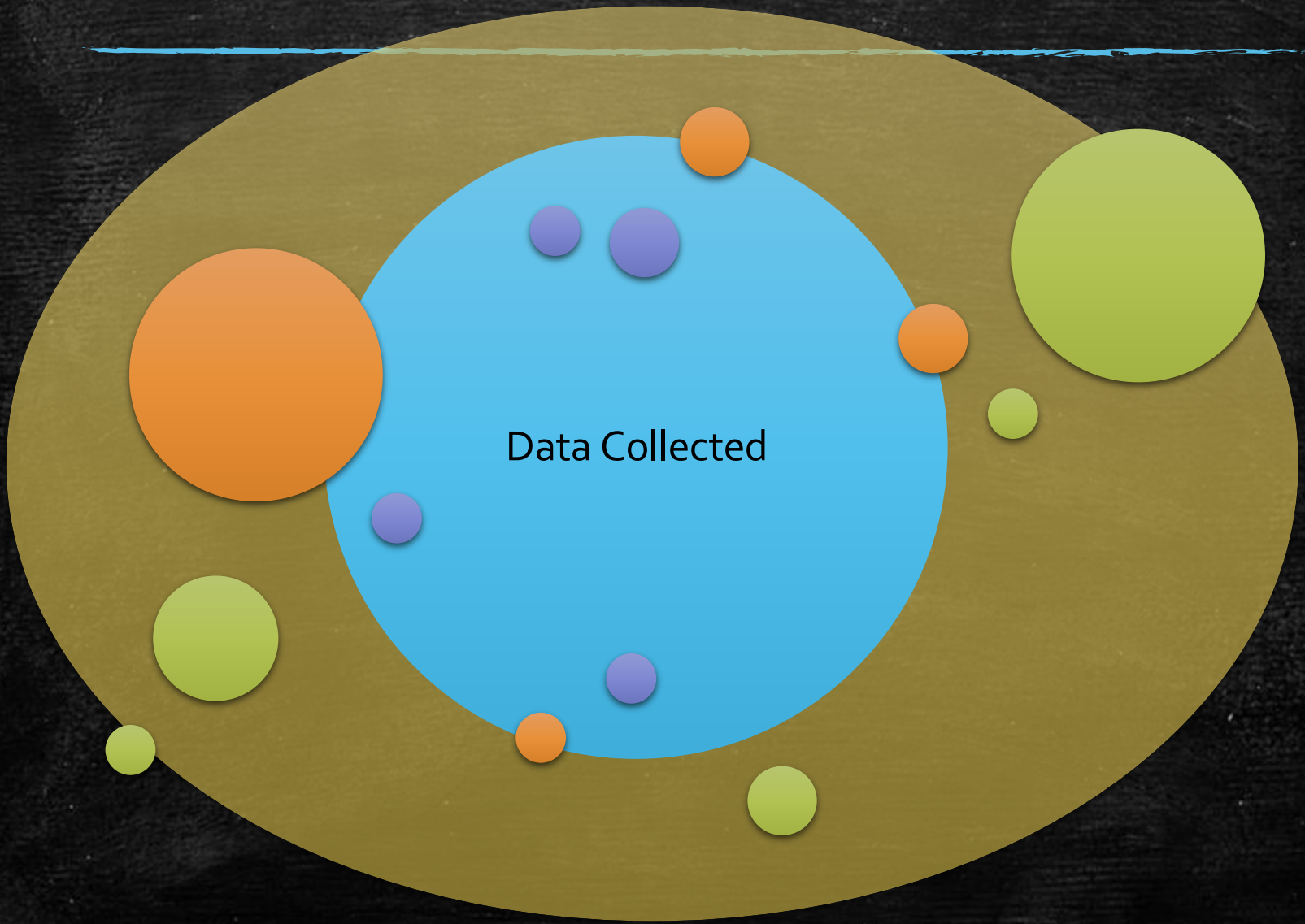
Census to the world: Show me the Data!

Roberto Rigobon, MIT


Big Distance!
Data
Information
Knowledge
The world is not lacking of Data
Lacking of Careful Empirics
Lacking of Managerial Data Analysis

Official Data Sources

New Technologies



Word of Caution

- Understand the Data Limitations
  - Identification
    - We need a Question
    - We need to understand the Process
  - Representativeness
    - We need to be cautious with our statements
- Different Data Sources
- Engagement and Research

Identification

- Who is more likely to drown while swimming in the ocean?
  - Actually, great swimmers!

- "Big Data" Approach:
  - Assume we collect every meet
    - Time, Style, Competitors, Place
    - Great measures of ability
  - Also assume we collect all the conditions in which the person drown

- What this data is missing is a measurement of the willingness of the individual to put itself at risk, given its ability

Identification

- The problem of identification cannot be solved by collecting more data, needs to
  - Collect the correct data
  - Design the test/survey properly - which requires understanding the process properly

- In many circumstances this cannot be solved with the data we observe.

Representativeness

- Social networks
  - Who has Facebook Accounts?
  - Who changes its status everyday?
  - Great data source with millions of observations from the same 200 teenagers.

- Reviews have the two problems
  - Decision to review is endogenous
  - Participation is non-representative

Data sources

- Not all data points have the same property
- Prices
  - Apple price of iPhone 6 on the web
  - Price of hotel in Hotel.com
  - Price of air ticket in Kayak.com
  - Prices on liquidation from a supermarket

Data Sources: Quality Ladder

- Utilities and Commercial activity
- Tracking (cellphones, cars, clicks)
- Web-scraping and Scanner from firm
- Measurement devices
- Emails and texts
- Reviews
- Social Networks

Engagement and Research

- Start with small pilots
  - Areas you already measure
  - Try to determine the improvement using alternative procedures to collect
    - Evaluate congruence
    - Evaluate timeliness
- Start with data sets that have the highest quality
- Need to change form of engagement
  - Openness and transparency to researchers
  - Data with a purpose:
    - Understand process
    - Understand weaknesses
    - Understand objective